Model Based Design of a Stereo Vision System for Intelligent Deep-Sea Operations

Model Based Design of a Stereo Vision System for Intelligent Deep-Sea Operations

Tomasz Łuczyński, Piotr Łuczyński, Lukas Pehle, Manfred Wirsum, Andreas Birk

PII: S0263-2241(19)30419-1
DOI: https://doi.org/10.1016/j.measurement.2019.05.004
Reference: MEASUR 6608

To appear in: *Measurement*

Received Date: 25 August 2018
Accepted Date: 4 May 2019

Please cite this article as: T. Łuczyński, P. Łuczyński, L. Pehle, M. Wirsum, A. Birk, Model Based Design of a Stereo Vision System for Intelligent Deep-Sea Operations, *Measurement* (2019), doi: https://doi.org/10.1016/j.measurement.2019.05.004

Model Based Design of a Stereo Vision System for Intelligent Deep-Sea Operations

Tomasz Łuczyński[a,c,*], Piotr Łuczyński[b], Lukas Pehle[b], Manfred Wirsum[b], Andreas Birk[a]

[a] *Robotics Group, Computer Science, Electrical Engineering, Jacobs University Bremen, Germany*
[b] *Institute for Power Plant Technology, Steam and Gas Turbines, RWTH Aachen University, Germany*
[c] *Institute of Sensors, Signals and Systems, School of EPS, Heriot-Watt University, Edinburgh, United Kingdom*

Abstract

A stereo vision system for deep-sea operations is presented. The system consists of cameras in pressure bottles, which are daisy-chained to a computer bottle. The system has substantial computation power for on-board stereo processing as well as for further computer vision methods to support autonomous intelligent functions, e.g., object recognition, navigation, mapping, inspection, and intervention. The model based design presented here includes two main aspects. First, a formalized approach to the component selection for the stereo set-up is introduced, i.e., given especially accuracy and baseline constraints as well as lens and imager options, an algorithmic analysis is provided. This approach is also of interest for the design of stereo systems in general. Second, the specific aspects of deep sea operations are addressed. This includes especially the validation and optimization of the pressure bottles for the cameras with a Finite Element Method (FEM). Experiments and results are presented, which include a validation of the stereo performance in air, robustness tests of the bottles in pressure tanks, and field trials of the complete system off the shore of Marseille on a commercial Remotely Operated Vehicle (ROV).

Keywords: Stereo Vision, Stereo Analysis, Finite Element Method (FEM), Finite Element Analysis (FEA), Deep Sea Operation, Intelligent Autonomous Systems, Remotely Operated Vehicle (ROV), Autonomous Underwater Vehicle (AUV)

1. Introduction

Cameras are essential sensors for marine systems as they provide detailed data at high update rates. Camera systems can accordingly be found on all kind of marine systems ranging from stationary platforms like landers to mobile robots like Remotely Operated Vehicles (ROV) or Autonomous Underwater Vehicles (AUV). When using more than one camera, depth information can be obtained via stereo processing. Mobile robotics in general can profit for many tasks like navigation, object recognition, or manipulation from this depth information from stereo vision [1]. This holds especially also for underwater applications [2, 3]. Unlike acoustic sensors in form of sonars, stereo vision provides high-resolution, relatively noise-free measurements at high update rates and at much lower cost. The caveat is of course that good visibility conditions are required. But this is typically the case for deep-sea operations where visibility is usually very good and the illumination conditions are perfectly controlled by artificial light.

We present in this article a model based design of a stereo vision system for on-board processing in deep sea operations. Model based design in engineering employs digital models and simulations to optimize parameters, to find and to correct potential flaws, and to test the limits of the product before it is even manufactured. This includes two main contributions in this article. First, a formalized approach to the component selection for the stereo set-up is presented, which is of interest for the design of stereo systems in general. Second, we address the specific aspect of deep sea operations, among others the validation and optimization of the pressure bottles for the cameras with a Finite Element Method (FEM).

As mentioned, our perspective on the analysis of the stereo system is from a component selection perspective, i.e., it is on determining the best choices among different hardware options under formal constraints in a formal algorithmic way. In a recent publication [4], some considerations towards choosing the cameras in the design stage of the system are also presented. However, the focus in [4] is on monocular visual odometry. In [5] an other attempt is made to formalize the process of designing stereo systems. However, [5] focuses on optimizing the coverage of the human activities space. For a 3D reconstruction, different factors must be considered as shown in the analysis later on.

There are also many publications on estimating the reconstruction errors that result from the quantization error in a stereo system [6, 7, 8, 9]. But a quantization error analysis alone simply leads to the conclusion that for the smallest errors, a baseline of infinite length should be used

*Corresponding author: *t.luczynski@jacobs-university.de*

[10], which is obviously impractical. Furthermore, it is also not an appropriate analysis when taking the curvature of the observed objects into account. In [11], an effort is made to define the optimal finite baseline based on extended analysis of the quantization error and the minimization of its depth component. However, this only leads to results where the optimal finite baseline can be specified for a given point in space, it can hence not be used during the design of a general purpose stereo vision system.

The second part of our model based design deals with the aspects of deep sea operations, especially the design and optimization of the housings for the stereo cameras. For this purpose, a FEM analysis is performed. A FEM is of course often used in a standard design process, especially in the area of structural analysis. The FEM can determine the deformation and stresses in the material. Consequently, the maximum permissible load applied to the investigated model can be determined based on material failure theory. Based on a specific use-case and the subsequent analysis of the mechanisms of stress development, the components with the highest stresses can be redesigned. Hence, a finite element analysis (FEA) can significantly accelerate the design process.

There are accordingly several examples of FEM for structural analysis in underwater applications. In [12], FEA is employed in the design process of a submarine radome. The numerical investigation allow the confirmation of the assumed safety limits and they indicate the critical zones in the radome, in which the highest stresses occur. In [13], an unmanned underwater vehicle is analyzed to operate at 6,000m water depth. Other examples of FEA in underwater applications include [14], [15] and [16].

But for the development of pressure bottles for marine systems, it is a standard practice to just use general design guidelines in combination with time-consuming laboratory testing. This is usually sufficient as pressure bottles tend to be quite simple in their design. The (relatively) low cost of standard bottles is also a factor, i.e., a failure in a pressure test in the lab destroys the bottle but it is a financially viable approach. Camera housings are different in this respect. They are more complex in their design and the camera windows tend to be very costly. A strong point of a FEM is its applicability to complex geometric shapes and loads, as it is the case for our camera housings. But the analysis performed for our system is computationally non-trivial. In the work presented here, the FEM is hence executed on a high-performance computer (HPC), namely the RWTH Bull Cluster in Aachen.

The rest of this article is structured as follows. Sec.2 introduces a formal approach to the selection of components and a corresponding accuracy analysis, which can be used for stereo systems in general. In Sec.3, this approach is used in a concrete application case, namely for a stereo system for deep sea operations. The system requires suited pressured bottles, which are analyzed by a FEM in Sec.4. Experiments and results are presented in Sec.5, which include a validation of the component selection approach in air, robustness tests of the system in a pressure tank, and field trials of the complete system off the shore of Marseille on a commercial Remotely Operated Vehicle (ROV). Sec.6 concludes the article.

2. Component Selection for Stereo Vision in General

2.1. Formulation of the Stereo Design Problem

When designing a stereo vision system, there are numerous decisions to be made, especially:

- Selection of the camera
- Selection of the lens
- Length of the baseline
- Vergence of the cameras

At the same time, these choices are limited by constraints:

- Minimum observing distance where the system should provide a 3D coverage
- Maximum length of the baseline that may be accommodated in the given application
- Desired accuracy of the 3D reconstruction

In the following, each aspect is shortly discussed. Camera selection has many aspects, e.g., cost, lead time, etc. When designing a stereo system, an interface for the data-transmission must also be chosen: USB, GigE, Fire Wire, etc. For stereo vision, especially when the system is mounted on a moving platform and/or there is dynamics in the observed scene, it is important that the data-interface of the cameras supports synchronization, i.e., it must be ensured that the two images of the cameras are taken at exactly the same moment in time. Other than that, the data-interface is mentioned here only for the sake of completeness. While each option has its advantages and disadvantages, overall this choice is not relevant for the design considerations discussed in this section. The only important factors for camera selection in this context are the resolution of the image sensor matrix, short imager, and the size of a single pixel.

Since the same imager may be mounted in different cameras, from now on the choice of the image sensor and not of the camera containing it will be discussed. There are some general rules that apply to this selection. A larger imager gives a bigger field of view for the same focal length of the lens. A larger pixel size results in higher sensitivity and better performance in low-light conditions. On the other hand, the larger the pixel, the larger the quantization error.

The lens must be compatible with the camera in terms of mounting, the size of the imager, etc. However, the most important factor from a formal viewpoint is its focal

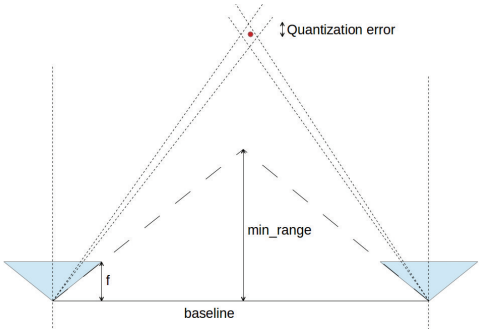


Figure 1: An illustration of the quantization error in a stereo camera system.

length. Increasing the focal length will decrease the quantization error at the cost of decreasing the field of view.

The length of the baseline and the vergence are important for the performance of the a stereo system. Increasing the baseline increases the accuracy of the 3D reconstruction by minimizing the quantization error. On the other hand, this also increases the minimum observation distance. This problem can be solved by verging the cameras. However, this narrows the field of view of the system. The baseline is also very often limited by other factors, e.g., the maximum length that can be accommodated on the system where the stereo camera is mounted.

2.2. The Minimum Range of a Stereo System

In the following, we assume that both cameras in the setup are identical, i.e., with the same lenses, and that they are symmetrically verged. Furthermore, the analysis is - as usual for cameras - restricted to the XZ plane. Following notations are used (Fig.2):

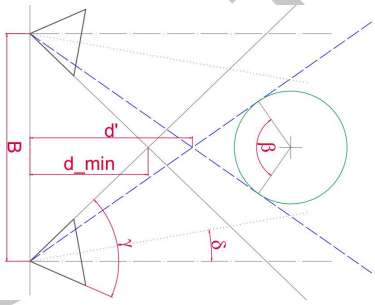


Figure 2: Overview of the general stereo setup considered here and the related notations.

- B is the length of the baseline
- f is the focal length of the lens
- γ is the field of view of the camera
- δ is the vergence angle
- $s.w$ is the width of the imager s in pixels

Figure 3: An example illustrating the influence of the radius of a cylindrical object on its perception with stereo vision. Here, three cameras are placed 25cm from each other, i.e., there is a 50cm baseline between the left and the right one, and 25cm between the center and the left, respectively right one. The perceived cylinders have 120mm, 80mm and 45mm diameter.

- $s.p$ is the size of a pixel of the imager s
- d_{min} is the theoretical minimum range of the system
- β is the angle defining the minimum arc of an object that should be observable
- d' is the minimum range corresponding to the β angle

Usually, the minimum range of a stereo system is defined by the border pixels of the two cameras (Fig. 1). However, this is only valid when observing a planar surface. To generalize this case, let us approximate the observed non-zero curvature of the object with a sphere or a cylinder, i.e., a circle in the projection on the XZ plane. When observing an according object, its surface reconstruction is only possible for an arc that can be seen by both cameras. Let us define this arc with an angle β. This redefines the minimum range to be d' (Fig.2).

The selection of β depends on the size of the objects that are expected to be perceived by the system. In the limit case, i.e., for a plane, the angle β is infinitely small. An illustration of the non-planar analysis using cylinders is shown in Fig.3. There, three cameras A, B and C are placed with 25cm between A and B, respectively B and C. Three cylinders with diameters of 120mm, 80mm and 45mm are placed in the center in front of them. It can be noticed that as the diameter gets smaller, the size of the surface patch that can be observed from two cameras also shrinks. But the angular measure of this patch remains constant. It must be noted that the minimum range d' derived in this way is shorter than the actual distance to the surface of the object. Nevertheless, it is defined here in this way, as it make the results of the algorithmic component selection more robust when using the stereo system in scenes with more complex shapes.

2.3. Non-Verged Case

First, the simplest case is analyzed, namely with no vergence ($\delta = 0$), which serves as basis for the selection algorithm. The case of no vergence leads to a field of view of the system that is identical to the field of view of the cameras. It is most versatile and hence for example often used in mobile robotics. In the following, the constraints within the analysis are discussed.

2.3.1. Maximum baseline constraint

There are two factors limiting the maximum baseline that can be used. The first is the maximum length allowed by the user B_{usr}, i.e., a design constraint due to limited space allocation for the system. The second, less trivial one, comes from the minimum range, respectively β constraint. Note that if the following analysis is to be invariant to the object size, the minimum β_{min} angle must be specified. Its value is determined by assuming that the whole object will be in the field of view and the observable part of it is defined by light rays tangential to its surface (Fig. 2).

$$\beta_{min} = 2 * (90 - \frac{\gamma}{2}) \tag{1}$$

If input $\beta < \beta_{min}$ is increased to the minimum value, this preliminary analysis can be used without specifying the size of the object. Having specified β, the maximum baseline can be calculated:

$$B' = \frac{2 * d'}{\tan \frac{\beta}{2}} \tag{2}$$

Finally, the first constraint is formulated:

$$B_{max} = min(B', B_{usr}) \tag{3}$$

2.3.2. Minimum baseline constraint

The minimum baseline constraint is derived from the minimum accuracy specified by the user. Even if we assume the calibration and the stereo processing to be perfect, an inherent quantization error always remains. Let us consider point P lying on a center line between the cameras at the distance d_a given by the user. To model the quantization error, we assume that the x coordinate of the projected points can vary by $\frac{s.p}{2}$. This constraint is iteratively handled. The baseline is increased by a fixed step and the accuracy is tested, i.e., it is checked if it already fulfills the given value. This way, minimum baselines B^i_{min} are determined (one for each specified distance/accuracy pair). The final constraint is:

$$B_{min} = max(B^1_{min}, ..., B^i_{min}) \tag{4}$$

As a result of this, the specified accuracy requirements are fulfilled.

2.3.3. Full Algorithm for the Non-Verged Case

Once both the minimum and the maximum baseline constraint are found, the solution can be derived. Three major cases can be identified. In the first case, $B_{min} > B_{usr}$, there is no solution. It means that even at the maximum baseline allowed by the user it is impossible to achieve the required accuracy. The second case is $B_{max} < B_{min} < B_{usr}$. This means that there is no solution when using two cameras, but adding more cameras will potentially help. A pair with a wider baseline will guarantee the required accuracy, and a shorter baseline will give the coverage of the objects close to the system. Of course, in some cases this may not be an option as a very high accuracy at short distances may be required. Finally, when $B_{min} < B_{max}$, a solution exists. In general, any length of the baseline within this range will satisfy the requirements specified by the user. However, it needs to be noted that additional sources of error may occur in both the quality and density of the reconstruction, so taking the border values of the returned range is not recommended.

The algorithm requires the following inputs: β, d', B_{usr}, the list of imaging sensors to be considered S, the list of available focal lengths F, the list of required accuracies at given distances (d^i, a^i). $NULL$ in Algorithm 1 represents the case with no solution.

2.4. Limited Vergence Case

The second case handles the situation when the user allows vergence, but prefers to maintain an unlimited maximum range. In other words $\delta < \frac{\gamma}{2}$. The equation for β_{min} (1) must then be modified:

$$\beta_{min} = 2 * (90 - \gamma) \tag{5}$$

A vergence angle δ must also be provided. If a solution can be determined, the vergence is calculated as:

$$\delta = 90 - \arctan(\frac{2 * d'}{B}) - \frac{\gamma}{2} \tag{6}$$

The changes in the Algorithm1 must be accordingly made to (6). Vergence δ is also limited to be positive - there is no point in turning the cameras away from each other. The results of the calculation of δ are given for B_{min} and B_{max}.

2.5. Unlimited Vergence Case

In the last possible scenario, the vergence δ is not limited. This may be useful when using lenses with a long focal length to model a closed space of known relative position with high accuracy. In this case, there is the risk that one of the distances for which the desired accuracy is specified is beyond the maximum range d_{max} of the system. Therefore, a new constraint for B_{max} must be formulated.

$$\gamma < \arctan(\frac{2 * d_{max}}{B_{max}}) - \arctan(\frac{2 * d'}{B_{max}}) \tag{7}$$

However, the value of B_{max} can not be directly derived from (7). Therefore, the Algorithm 1 is modified by adding Algorithm 2 to iteratively solve this constraint. A simple solver with a fixed step-width is proposed here, as this is only used for specifying initial cues for the further design decisions and an exact solution is not required.

Algorithm 1: Stereo system design algorithm (non verged case).

for *each* $s \in S$ **do**
  for *each* $f \in F$ **do**
    $f_p = \frac{f}{s.p}$
    $l = [-s.w; 0; f_p]$
    $r = [s.w; 0; f_p]$
    $l.normalize(), r.normalize()$
    $\gamma = \arccos(l^T * r)$
    if $90 - \frac{\gamma}{2} > \frac{\beta}{2}$ **then**
      $\beta = 2 * (90 - \frac{\gamma}{2})$
    end
    $B_{max} = \frac{2*d'}{\tan(\frac{\beta}{2})}$
    for *each* (d^i, a^i) **do**
      $B = 0.02$
      $step = 0.005$
      while *not_solved* **do**
        $B+ = step$
        $P = [-\frac{B}{2}; 0; d^i]$
        $Img = \frac{P*f}{P.z}$
        $Img_{min} = Img_{max} = Img$
        $Img_{min}.x = Img.x + \frac{s.p}{2}$
        $Img_{max}.x = Img.x - \frac{s.p}{2}$
        $P' = \frac{Img_{min}}{Img_{min}.x} * \frac{B}{2}$
        $P'' = \frac{Img_{max}}{Img_{max}.x} * \frac{B}{2}$
        $e = |P'.z - P".z|/2$
        if $e < a^i$ **then**
          solved
          $B^i_{min} = B$
        end
        else if $B < B_{usr}$ **then**
          solved
          $B^i_{min} = NULL$
        end
      end
    end
    if $max(B^1_{min}, ..., B^i_{min}) == NULL$ **then**
      $Results(s, f) = NULL$
    end
    else if $max(B^1_{min}, ..., B^i_{min}) > B_{max}$ &&$min(B^1_{min}, ..., B^i_{min}) > B_{max}$ **then**
      $Results(s, f) = NULL$
    end
    else if $max(B^1_{min}, ..., B^i_{min}) > B_{max}$ &&$min(B^1_{min}, ..., B^i_{min}) < B_{max}$ **then**
      $Results(s, f) =$ $MORE\_CAMERAS\_NEEDED$
    end
    else
      $Results(s, f) =< B_{min}, B_{max} >$
    end
  end
end
output: *Results*

Algorithm 2: Additional constraint for B_{max} (unlimited vergence).

$\alpha = \arctan(\frac{2*d_{max}}{B_{max}}) - \arctan(\frac{2*d'}{B_{max}})$
while $\alpha > \gamma$ **do**
  $B_{max} - = 0.05$
  $\gamma < \arctan(\frac{2*d_{max}}{B_{max}}) - \arctan(\frac{2*d'}{B_{max}})$
end

2.6. Setup Analysis after Component Selection

The algorithm presented above gives the range of baselines for which the given conditions are fulfilled: choosing a baseline closer to the maximum value increases the accuracy, choosing a shorter baseline increases the observable part of the surface (i.e., β). Both values are subject to additional sources of error. The accuracy discussed above provides a baseline, best case analysis; it can be lowered by errors in the camera calibration, motion blur, limits in the stereo processing, etc. The β angle may decrease due to the same factors, but also due to featureless textures and the quantization error. In this case, the quantization error will manifest itself in our example on the side of the cylinder, where both the background and part of the cylinder are visible within one pixel.

Some of these effects depend on the size of the object, therefore a second step in the design process is introduced now. After selecting the imager, the lens, the baseline, etc., these values are used to calculate the predicted values of the reconstruction accuracy and the β angle. This time β is calculated together with the predicted influence of the quantization error. Since this influence depends on the cylinder's diameter, it is presented in a form of a graph. The predicted value of β corrected for the quantization error (denoted as β_c) can be calculated as:

$$\sigma = \arcsin(\cos(\alpha) - \cos(\alpha)\tan\alpha\tan(\frac{\beta}{2}) - \frac{B\tan\alpha\cos\alpha}{2R\cos(\frac{\beta}{2})}) \tag{8}$$

$$\beta_c = \beta - 2(90 - \alpha - \sigma) \tag{9}$$

where α is the angle defined by the size of single pixel for a given focal length f and σ is just an auxiliary angle used in the calculation.

The accuracy of the system is calculated by estimating the maximum quantization error according to Algorithm 1. The results of the system analysis can then be presented in form of graphs (e.g., Fig.6).

3. Component Selection for the Jacobs Deep Sea Stereo System

3.1. Background and Motivation

As mentioned in the introduction, cameras in general and stereo systems in particular are important sensors for

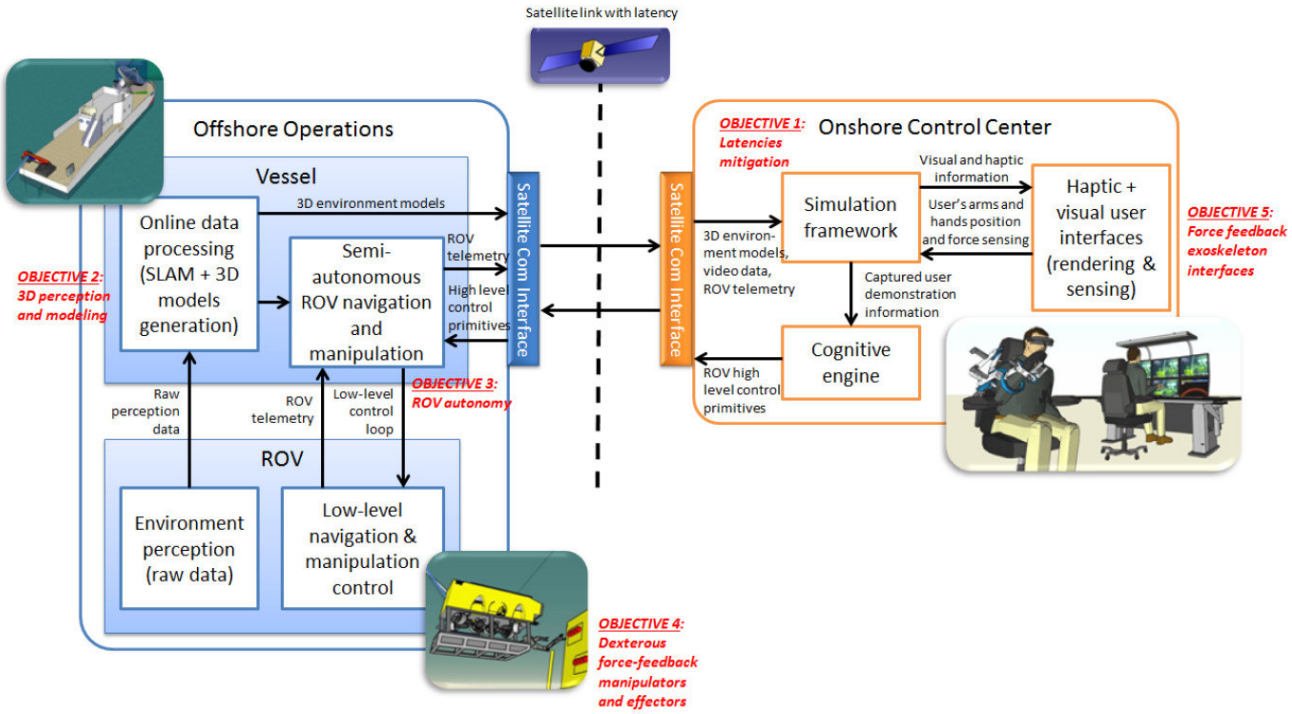


Figure 4: Within the DexROV project, underwater manipulation from an onshore control center is investigated to reduce operation costs of ROVs and to ease the involvement of mission experts. This requires increased autonomous intelligent functions on the ROV to compensate for latency and bandwidth constraints of the satellite link from the onshore center to the offshore support vessel. The stereo system presented here is the main sensor for this purpose.

marine systems. Our work presented here is motivated by the EU-project "Effective dexterous ROV operations in the presence of communication latencies (DexROV)". Commercial underwater operations are dominated by ROV, which require a highly qualified offshore crew, respectively even multiple crews working in shifts, who need to engage in a significant amount of low-level, manual control.

DexROV attempts to reduce the amount of robot operators required offshore by facilitating operations from an onshore control center and reducing the gap between low-level tele-operation and full autonomy (Fig. 4). The basic idea is that the user interacts with a real time simulation environment, and a cognitive engine analyzes the user's control requests and turns them into motion primitives that the ROV autonomously executes in the real environment - independently of communication latencies from a satellite connection [17].

State of the art ROV like the Sub-Atlantic Apache vehicle used in DexROV often rely on analog camera systems. The image quality of standard analog cameras is strongly influenced by the connection from the ROV to the support vessel. The main commercial ROV application areas, e.g., operations in oil- and gas-production, are in depths that are inaccessible to divers, i.e., in the deep sea. Due to the long tethers required for deep-sea operations, images are often noisy and there can be interrupts in the streams, which makes processing of the data by computer vision on the vessel quite challenging.

The use of digital cameras has therefore in general multiple advantages, especially when they are combined with computation power on-board of the vehicle to generate an intelligent vision system. First of all, an intelligent underwater vision system can minimize the data traffic over the tether from the ROV to the vessel. It allows an online adaptation of the image resolution, the compression factor, and the frame rate to optimally use the available bandwidth for the current task at hand. Furthermore, computer vision can be used directly on-board of the ROV to assist core capabilities up to the provision of autonomous functions. This processing onboard of the ROV minimizes latencies and increases robustness compared to processing on the vessel, which requires the transmission of the data over the limited data-connection of a tether.

3.2. Selection of the Cameras, Lenses, and the Baseline

In the following, the component selection approach presented in Sec.2 is used for the design of the Jacobs stereo system for DexROV. The list of imaging sensors considered for the system is presented in Table 1.

Given the specific use-case, the focal length of the lens is restricted to a maximum of 5.5mm to get a wide field of view. Furthermore, a non verged design is chosen. Since a flat panel glass window is used in the housing, the focal lengths of the lenses when being used underwater can be

Table 1: List of sensor models and most important parameters

Model	Resolution	Pixel size [nm]
Sony ICX414 CCD	648 x 488	9.9
Sony ICX274 CCD	1624 x 1224	4.4
Sony ICX818 CCD	1928 x 1448	3.69
Sony ICX267 CCD	1384 x 1032	4.65
Sony ICX687 CCD	1928 x 1448	3.69
Sony ICX674 CCD	1932 x 1452	4.54
Sony ICX694 CCD	2736 x 2192	4.54
Sony ICX808 CCD	2016 x 2016	3.1
Sony ICX285 CCD	1384 x 1036	6.45
Sony ICX625 CCD	2448 x 2048	3.45
Sony ICX825 CCD	1384 x 1032	6.45

estimated by increasing their nominal values by 33%. This is only an approximation, but it is sufficient for the purpose of component selection where only coarse options for imager and lens parameters are available. As presented in more detail later on in Sec.4.1, this is not sufficient when using the implemented system as a precise instrument, i.e., a more accurate camera model is used for the calibration and the image rectification once the components are determined.

Two accuracy checkpoints are specified. At 5m distance, the maximum error is set to be no more than 45cm. At 2m distance, it is set to be no more than 6cm. To be able to use the system among others for autonomous manipulation tasks, the minimum range is set to be 30cm. The maximum baseline is limited by the available space on the ROV where the system is integrated; it must be restricted to a maximum of 0.5m. The expected minimum value of β is set to 120°. The results of the corresponding analysis are summarized in the Table 2.

Based on these results, a Sony ICX285 CCD imager, respectively a Pointgrey Grasshopper2 1394b camera containing this sensor, with a 4.8mm focal length lens is chosen. The baseline is set to $0.3m$. To complete this selection, this setup is analyzed using the approach described before in Sec.2.6. The according results are shown in Fig.6.

3.3. Camera Interfaces and On-board Computer

For the sake of completeness, we shortly describe here also the other components of our underwater stereo system. An important aspect in this respect are the Firewire (1394b) interfaces of the cameras, which can not be implemented with standard underwater cables due to their electrical high-frequency (HF) demands. Note that this aspect holds for any data-interface of high-resolution cameras. Therefore, special underwater cables are used, which are designed for Ethernet applications and which are hence suited for the transmission of HF-signals at Gigabit/s. Note that Firewire has 9-pin connectors but only 8 lines

Table 2: Table of results given by the component selection algorithm

Focal lengths / Sensors	2.8 mm	3.8 mm	4 mm	4.8 mm	5.5 mm
Sony ICX414 CCD	More cameras required	More cameras required	0.335 /0.34641 m	0.28 /0.34641 m	0.245 /0.34641 m
Sony ICX274 CCD	0.21 /0.34641 m	0.155 /0.34641 m	0.15 /0.34641 m	0.125 /0.34641 m	0.11 /0.34641 m
Sony ICX818 CCD	0.18 /0.34641 m	0.13 /0.34641 m	0.125 /0.34641 m	0.105 /0.34641 m	0.09 /0.34641 m
Sony ICX267 CCD	0.225 /0.34641 m	0.165 /0.34641 m	0.16 /0.34641 m	0.13 /0.34641 m	0.115 /0.34641 m
Sony ICX687 CCD	0.18 /0.34641 m	0.13 /0.34641 m	0.125 /0.34641 m	0.105 /0.34641 m	0.09 /0.34641 m
Sony ICX674 CCD	0.22 /0.34641 m	0.16 /0.34641 m	0.155 /0.34641 m	0.13 /0.34641 m	0.115 /0.34641 m
Sony ICX694 CCD	0.22 /0.34641 m	0.16 /0.34641 m	0.155 /0.34641 m	0.13 /0.34641 m	0.115 /0.34641 m
Sony ICX808 CCD	0.15 /0.34641 m	0.11 /0.34641 m	0.105 /0.34641 m	0.09 /0.34641 m	0.08 /0.34089 m
Sony ICX285 CCD	0.31 /0.34641 m	0.23 /0.34641 m	0.22 /0.34641 m	0.18 /0.34641 m	0.16 /0.34641 m
Sony ICX625 CCD	0.165 /0.34641 m	0.125 /0.34641 m	0.12 /0.34641 m	0.1 /0.34641 m	0.085 /0.34641 m
Sony ICX825 CCD	0.31 /0.34641 m	0.23 /0.34641 m	0.22 /0.34641 m	0.18 /0.34641 m	0.16 /0.34641 m

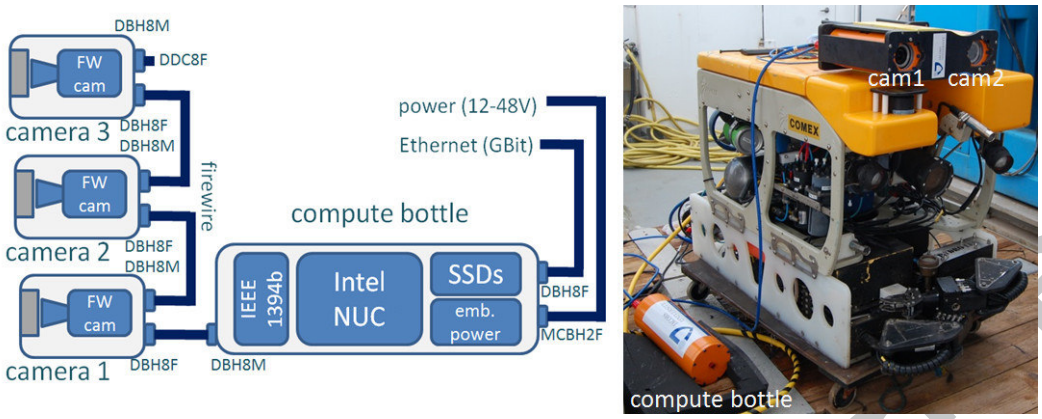


Figure 5: The camera system consists of a compute bottle for online processing power to which multiple cameras in pressure housings can be daisy-chained (left), e.g., as a stereo set-up with two cameras (right).

are used for data and power, i.e., it can be mapped to the 8 lines of Ethernet.

Concretely, Subconn underwater connectors and cables are used in form of female DBH8F, respectively male DBH8M male bulkheads at the bottles and D-P4TP24 HF-cables with a male DIL8M, respectively female DIL8F connector at each end. Due to the alternating genders at the two ends of each cable, each camera bottle has a well defined output (female), respectively input (male) for the daisy-chaining of the cameras to the compute bottle with an accordingly also male input. The last camera in the daisy-chain has obviously no input; its according bulkhead connector is hence sealed with a female DDC8F dummy.

The daisy-chaining allows to use also more than two cameras, e.g., to generate multiple stereo systems with different baselines. Three cameras C_1, C_2, and C_3 for example can form three stereo systems with the baselines $\overline{C_1, C_2}$ (e.g., short), $\overline{C_2, C_3}$ (e.g., medium), and $\overline{C_1, C_3}$ (e.g., long baseline). Due to payload and space constraints on the Apache ROV used in our application, only two cameras are employed in the use-case presented here.

An other core hardware element of the system is the computer for the stereo and further vision processing onboard of the platform where the system is integrated on. The computer used for this purpose is a Intel NUC with a 4th-Gen Intel Core i5-4250U, which provides reasonable processing power for this purpose. The i5-4250U CPU features a dual-core processor running at 1.3Ghz. Furthermore, the NUC is equipped with 16GB RAM and a 1TB HDD. The thermal design power (TDP) of the NUC is 15W. A DCDC-NUC with a wide-range input of 6-48V is used as power-supply. A mini-PCIe card is used to provide the Firewire interface.

The pressure bottle, in which the computer and its power system is integrated, is based - just like the camera bottles - on hardened aluminum alloy of type AlMgSi1 (EN AW-6082-T6), which transports heat very well with a thermal conductivity coefficient of 180 W/mK. Not only the cameras, but also the computer bottle can hence run under heavy computation load for unlimited time in-air, i.e., even without the additional cooling effects of underwater operations. In-air operations of the system including both testing and even software development were done over multiple days without ever encountering any problems from overheating.

Figure 6: Results of the stereo system analysis for a Sony ICX285 CCD sensor, 4.8mm lens, $B = 0.3m$, $d' = 0.3m$, and no vergence.

4. Camera Bottle Analysis with a Finite Element Method (FEM)

4.1. The Camera Bottles

Figure 7: One of camera housings in form of a pressure bottle with a flat-pane sapphire glass window.

One challenging aspect for deep sea operations is the enormous water pressure that the equipment has to withstand. This holds especially for the camera housings as the pressure bottle has to include a window part, behind which the camera is placed. Furthermore, it is standard practice for some deep sea components, e.g., for manipulators and often also for computer or other electronics bottles, to completely fill them with oil, which is unlike air incompressible. This is not an option for camera housings. This would require a replacement of any bit of air in the system with oil, including also the air-filled spaces in between the camera lenses as well as in the space in between the imager and the lenses. But the optical properties of a camera and its lens(es) rely on the refraction of air to glass and vice versa. Even if a suited, transparent oil would be available, the refraction indices of oil and glass are too close to each other to be usable for this purpose.

For our camera housing, a flat panel sapphire glass window is used with 20mm thickness (Fig.7). While dome ports provide in theory optical advantages, they have to be specially designed and manufactured to fit the camera. The engineering requirements of the manufacturing process typically lead to solutions based on plexiglass, which is inferior to sapphire glass in terms of optical properties and handling; plexiglass is, e.g., easily scratched during missions and even during system maintenance. Furthermore, the required design and especially the manufacturing processes typically lead to significantly higher costs for domes than for flat pane windows. Last but not least, domes require an exact placement of the camera in their optical center, which makes assembly, operation, and maintenance more challenging when precisely focused data is required.

On the other hand, flat ports introduce significant distortions due to the refraction at the air-glass and glass-water interfaces [18, 19, 20, 21]. But as shown in recent work [22], these effects can be easily compensated by calibration and rectification using the pinax camera model. This model is derived from the axial camera model, which is physically correct for flat pane interfaces but difficult to tackle in practice as points on its projection plane do not

lay on straights ray from the scene. Using realistic constraints on the placement of the camera to the window, namely by assuming short distances, the projection function of the axial model is combined in pinax with a virtual pinhole model.

The pinax model allows the pre-computation of a lookup-table for very fast refraction correction of the flat-pane with high accuracy. In doing so, the pinax model takes the refraction indices of water into account, especially with respect to its salinity, for which nominal values for fresh and salt water are sufficient. Therefore, it is sufficient to calibrate the camera only once in air. As shown in detail in [22], the model is not only much more convenient to use then state of the art methods that require in-situ calibration, it is also more accurate and leads to high quality rectification required for, e.g., stereo processing.

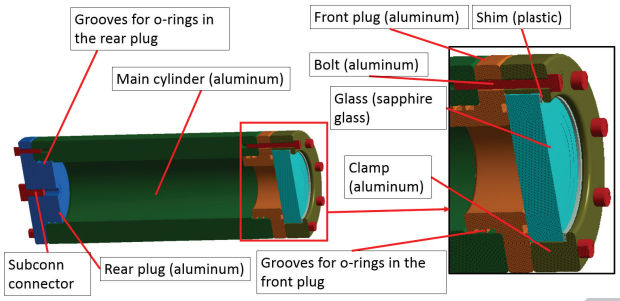


Figure 8: Components of the camera housing.

The camera bottle is sealed with a standard double o-ring design for the rear cap and the front cap with the window mount (Fig.8). Furthermore, all contact surfaces, e.g., the flanges of the caps and the ends of the pressure cylinder are machined to a glossy finish (Ra 0.4). In combination with the high pressures of deep sea operations, this has an additional sealing effect. The Ra 0.4 finish also provides excellent contact and hence mitigates stress in the material.

The plastic shim in front of the camera (Fig.8) is only used to prevent the aluminium clamp from scratching the glass; it has no sealing effect. When submerged, the water presses the window against the front cap and its o-ring. At this stage, both the clamp in front of the window and the plastic shim bear no load. Last but not least, the edges of the glass windows are chamfered to minimize the risk of cracks from stress at the edges.

4.2. Overview of the FEM Analysis

The main goal of the FEM analysis presented below is the determination of stresses and deformations in the camera housing for predefined pressure loads. In the following, the main modeling steps and the boundary conditions are shortly discussed. The main components of camera housing are shown in Fig.8. The main cylinder has a length of 265mm and an inner diameter of 95mm, which are determined by the dimensions of the camera, its lens and the cabling from the bulkhead connectors at the rear cap. Its outer diameter is 120mm, which is determined, respectively validated by the FEM presented below. The design parameters of the rear cap and of the front cap with the window mount are determined via the iterative FEM validation and optimization.

The rear/front caps and the window clamp are attached to the main cylinder by a threaded bolt connection, which in the numerical FEM analysis is modeled as a bonded contact. On the other contact surfaces of the model, appropriate friction coefficients are defined. Based on the geometry, a fine, mixed computational mesh of hexahedral and tetrahedral elements is generated. Due to the high resolution of the mesh, the FEM simulations are computationally non-trivial. Therefore, a high performance computer, namely the RWTH Bull Cluster in Aachen, is used for this purpose. The relevant material properties are listed in Table 3.

The main design considerations are presented below. As mentioned, the FEA also includes the optimization of other design parameters, e.g., of dimensions of the front and the rear cap. First, the linear elastic FEM simulations are presented for the pressure load at a water depth of 1,500m, which is the expected maximum operation depth of 1,200m of our use-case plus a nominal safety margin of 25%. Second, the maximum theoretical operating depth of the camera housing is discussed, as the system can be of interest for further applications and to determine the real safety margin for the current use-case. Individual components of the housing in which plastic deformations occur are subsequently recalculated using the elastic-plastic FEM approach.

The value of pressure acting on the submerged surfaces of the housing depend on the water depth. According to Pascals principle, the hydrostatic pressure p_{load} at water depth h is defined as:

$$p_{load} = p_0 + \rho g h \tag{10}$$

where p_0 stands for the atmospheric pressure, ρ for the (nominal) density of water and g the gravitational acceleration. The housing is filled entirely with air at atmospheric pressure of about $0.1MPa$. Additionally, it is assumed that, due to the high ambient pressure, the contact surfaces between individual components of the housing perfectly adhere to each other. The pressure distribution applied to the inner and outer surfaces of the investigated numerical model is shown in Fig.9.

4.3. FEM analysis for 1,500msw

First, we present the stress distribution in the camera housing during operation at 1,500m depth, or more precisely 1,500 meter sea water (msw) as the nominal unit of pressure used for marine systems, which is defined as $1msw = 0.1bar = 10KPa$. Accordingly, linear-elastic FEM simulations with a pressure load of $p_{load} = 15MPa$

Table 3: Material properties for the camera housing.

Parameter / Material	Young's Module [MPa]	Poisson's Ratio [-]	Tensile Yield Strength [MPa]	Tensile Ultimate Strength [MPa]
Alum. EN AW 6082	70,000	0.33	260	310
Sapphire Glass	345,000	0.34828	-	340

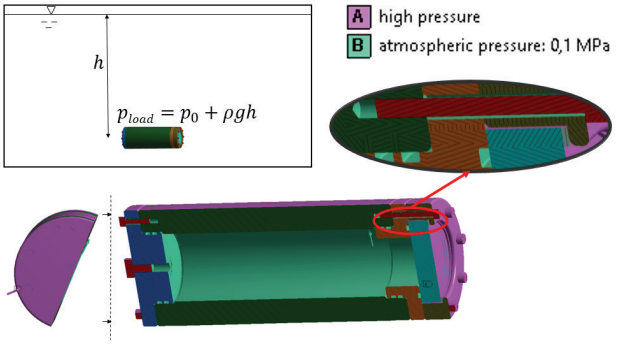


Figure 9: Pressure distribution on the inner and outer surfaces of the camera housing.

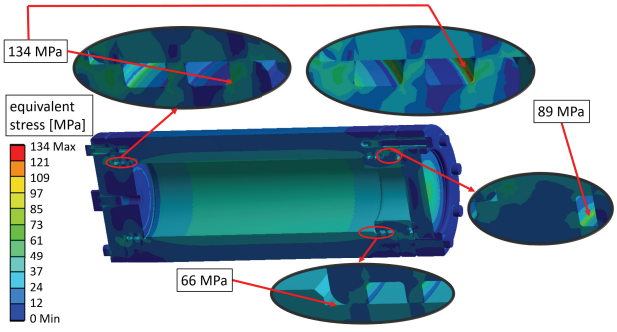


Figure 10: Stress distribution in the camera housing for 1,500msw, i.e., $p_{load} = 15MPa$.

are conducted. The resulting values of equivalent stress (von Mises stress) are shown in Fig.10.

The highest stresses occur in the grooves designed for the sealing o-rings. The second highest stress is located in the main cylinder, near the end of the contact to the front cap. All the high stresses are significantly below the yield strength of the hardened aluminium (260 MPa, cp. Table 3). Hence, no plastic deformations in the camera housing occur in a water depth of 1,500m, and the operational safety of the underwater vision system for its planned missions is confirmed.

4.4. Determination of the Maximum Depth

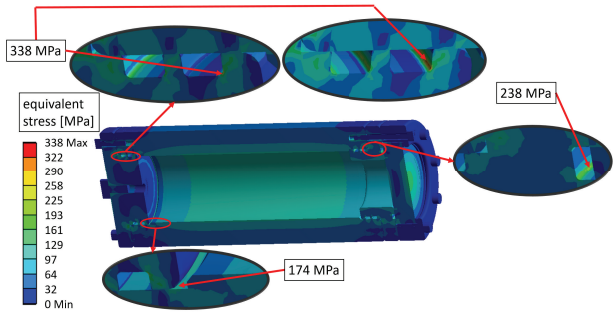


Figure 11: Stress distribution in the camera housing at 4,000msw, i.e., $p_{load} = 40MPa$.

To determine the maximum water depth in which the camera housing can safely operate, the values of strains and stresses are furthermore calculated for different pressure loads p_{load}. The pressure load leading to the first local plastic deformations is considered to be the operation limit.

The FEA for different depths shows a p_{load} equal to 40MPa (4,000msw) as a maximum permissible load. The results of the linear elastic calculation are shown in Fig.11. The highest local stress arises in the groove of the rear plug and its value exceeds the yield strength of the aluminium (260 MPa, cp. Table IV). Hence, in the next step of the analysis, the rear cap is modeled by means of an ideal elastic-plastic FEM. The fillets of the groove in which the highest stresses occur are re-meshed using very fine hexahedral elements in order to exclude the singularities in the FEM sub-model of the rear cap.

As a result of this FEA, the distribution of plastic deformations in the rear cap is obtained, with the highest local value of equivalent plastic strain equal to 0.12%. The local plastic deformations do not propagate deeper into the material and crack development is hence not to be expected. Similar conclusions can be drawn from the original linear-elastic FEA.

According to the recommendations derived in [23], the higher values of local stresses obtained by means of linear-elastic FEA should be averaged for several neighboring elements (depending on mesh spacing). The resulting values

of averaged equivalent local stresses in the camera housing do not exceed the specified value of the yield strength of the aluminum. Nevertheless, the pressure load of 40 MPa should be considered as a theoretical operational limit of the camera housing. Higher values of pressure load will unavoidably lead to higher plastic deformations, increase the risk of crack development and, in extreme cases, may result in the failure of the material.

4.5. Failure Mitigation for the Maximum Depth

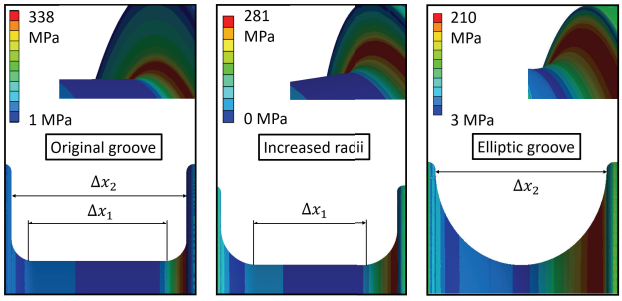


Figure 12: Possible adaptations of the o-ring groove geometry ($p_{load} = 40MPa$)

According to the results discussed above, the highest stresses for the 15 MPa and the 40 MPa pressure loads occur in the groove of the rear cap (cp. Fig.10 and Fig.11). The high values of local stresses in the groove are caused by the notch effect and the small radii of the fillets. Thus, in order to increase the maximum permissible operating depth, respectively to prevent plastic deformations and stress in long-time operations, minor changes are introduced in the design of the rear cap.

In Figure 12, two possible adaptations of the groove geometry are presented. In the first case (called "Increase radii"), the radii of the fillet are simply increased without changing the distance Δx_1. In the second case (called "Elliptic groove"), the distance Δx_2 is held constant as the elliptic shape is applied to the groove geometry.

The linear-elastic FEA reveals that for the pressure load of 40MPa, the two alternative modifications in the housing design result in a significant reduction of the maximum values of the von Mises stress, namely by 17% (Case: Increased radii, Figure 12), respectively even 38% (Case: Elliptic groove, Figure 12). In the elliptical shaped groove, no plastic deformations are to be expected as the tensile yield strength of the aluminium (260 MPa, cp. Table 3) is not exceeded.

5. Experiments and Results

5.1. In-Air Evaluation of the Stereo Component Selection

First, the selection of the specific components for our deep sea stereo system (Sec.3.2) are evaluated in in-air experiments. The experiment also serves as illustration for the validity of the underlying formal framework (Sec.2).

In this experiment, the selected cameras and lenses are used to set-up three different stereo systems. The first set-up uses a baseline $B = 0.25m$, which fulfills all assumptions. The second system has a baseline shorter than the range determined by the selection algorithm (Alg.1), namely $B = 0.1m$. Finally, the last system has a baseline that is longer than the suggested range for a suited baseline, namely $B = 0.4m$. For each system, a formal accuracy analysis is performed, as described in Sec.2.6.

The stereo capabilities are then empirically checked for each system. A cylinder is placed in front of each set-up (Fig.13), with lines marked on its surface every 10 degrees to simplify the measurement of the β angle, i.e., the predicted values can be compared to the data in the real-world set-ups. It is of course important to note, that the actual scene reconstruction capabilities of the stereo processing also depend on many other factors, e.g., object/scene textures, light conditions, reconstruction method, etc. The results of the comparison are presented in Table 4.

Table 4: Predicted values of β_p and measured β_m for three stereo set-ups.

Baseline / Distance	0.1m	0.25m	0.4m
0.3m	$\beta_p = 143$ $\beta_m = 135$	$\beta_p = 120$ $\beta_m = 115$	$\beta_p = 98$ $\beta_m = 95$
0.5m	$\beta_p = 146$ $\beta_m = 140$	$\beta_p = 133$ $\beta_m = 125$	$\beta_p = 118$ $\beta_m = 115$

To minimize the errors coming from the aforementioned sources of reconstruction errors, the β angle of the arc visible on both cameras is estimated manually with the help of the lines printed on the cylinder. This estimation is also given up to 5 degrees accuracy. As expected, the shorter the baseline, the bigger the part of the object that is seen by both cameras at the same time. Measured values are only slightly smaller than predicted, indicating that the component selection algorithm works indeed as intended.

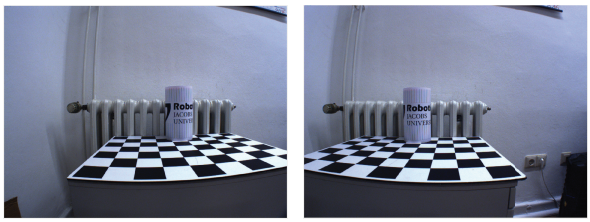


Figure 13: Example views from the left and the right camera on the cylinder used in evaluation process.

In a second experiment, the accuracy prediction for the system is evaluated. Again, the three set-ups with a $10cm$, $25cm$, respectively $40cm$ baseline are used. Each set-up observes a calibration pattern at $0.5m$, $1m$ and $2m$ distances. The markers on the calibration pattern can be

detected by machine vision with subpixel accuracy, hence giving a good ground truth reference. The positions of the markers are then rounded to full pixels as input for the stereo triangulation, and their accordingly determined 3D positions are compared to the ground truth. This reveals the effect of the quantization error when reconstructing the position of an observable feature in the environment. The results of this experiment are presented in Fig.14. Since our accuracy algorithm predicts the maximum error coming from the quantization, it is also compared to the maximum error in the experiment.

Figure 14: Evaluation of the accuracy prediction algorithm.

The results shown that the measured values correspond to the predicted ones quite well. As the distance to the object grows and the maximum quantization error becomes very large, the measured error becomes slightly smaller. This is simply due to the fact that the worst possible scenario did not occur. This indicates that the accuracy prediction step works as intended.

5.2. Bottle Tests in a Pressure Tank

The failure of pressure bottles is not only disastrous for the components they contain. It can also have catastrophic effects on the platform they are integrated on due to the shock-wave of the implosion and due to the change in buoyancy of the overall system. Hence, it is good practice - if not even mandatory for legal and insurance reasons - to test new marine systems in pressure tanks before they are used in the field.

Accordingly, the pressure test facilities of the Center for Marine Environmental Sciences (MARUM) in Bremen were used to validate our stereo system for deep sea operations. Fig.15 shows the result of this test. The system is submerged in the test tank and the pressure is increased to 150 bar, i.e., the conditions at 1,500msw, which correspond to the maximum depth of the planned use-case plus 25% safety margin. The pressure is maintained for an hour, which is the standard time considered to be more than sufficient to detect design or material flaws in the bottles, i.e., to cause leakages or an implosion. The pressure is then slowly decreased again.

Figure 15: Pressure profile during system validation in a test tank.

The real-world pressure tests confirm the results of the FEM analysis (Sec.4), i.e., the system withstands 1,500msw without any problems. Note that the standard design approach for marine systems often includes multiple iterations of pressure tests, i.e., the pressure bottles are first designed according to general guidelines and best practices, then tested, and (possibly multiple times) adapted and corrected if the pressure tests fail. For our camera housings, the failure of a bottle in pressure tests would not only be inconvenient as it would require an additional design and production cycle, it would also be very costly due to the high-end sapphire glass windows. The presented FEM analysis prevents - or at least reduces the risk of - costly failures in pressure tests.

5.3. Field Trials

Finally, results for the underwater performance are presented. The stereo system was tested in extensive field trials in the Mediterranean Sea offshore of Marseille in June/July 2017 and in June/July 2018. Each test campaign lasted for two weeks with multiple test dives. The results presented below are from a test dive in 30m depth with ambient light conditions. The baseline of the stereo system was set to 30cm in this trial. To validate the

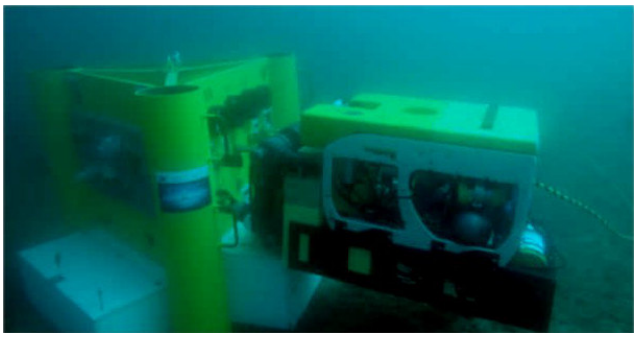


Figure 16: The Apache ROV with the stereo system and the DexROV panel in an underwater experiment in the Mediterranean Sea.

overall DexROV scenario, a test panel was developed by the project partner "Compagnie maritime d'expertises (COMEX)", that also provided the Apache ROV and the Janus-II vessel for the sea trials (Fig.18). The panel emulates different application scenarios, e.g., oil&gas interfaces based on the ISO 13628 standard including valves and wet-mate connectors.

As already discussed in Sec.3.1, our stereo system is the main sensor for a variety of tasks [17], including navigation, mapping, object recognition, and intervention. Fig.17 shows a few qualitative results from the wide range of research topics on underwater vision related to this [24, 25, 26, 27]:

(a) Image enhancement is used to improve the performance of higher-level methods, e.g., of stereo processing and marker detection.

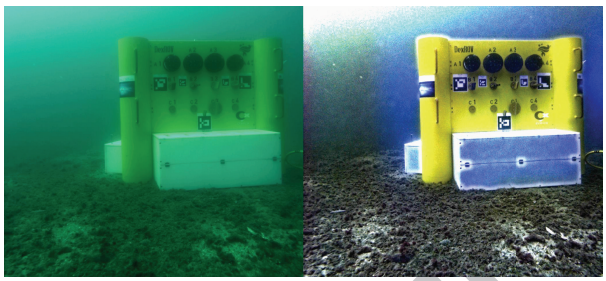
(b) While stereo processing with the standard SAD-algorithm leads to good results for the natural environment (see also (e) on the left), it suffers from the well-known issues when there are featureless regions like the uniform parts of the panel.

(c) A stereo algorithm on the basis of a Deep Learning Neural Network is hence used for getting dense data of the panel.

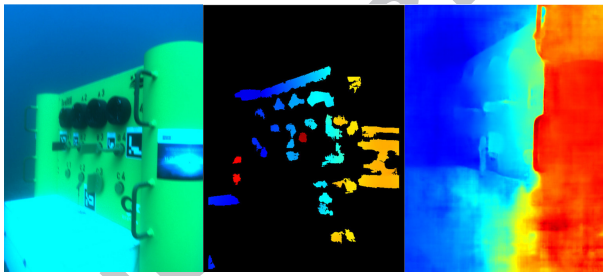
(d) The localization and the estimation of the state of valve-handles on the panel, especially their orientation, is determined by an active contours method with super-ellipse fitting in combination with a Hough transform.

(e) Depending on the task, different forms of maps can be used, e.g., colored point clouds for the natural environment, dense range information for the panel and its elements, or an octree that is well suited for obstacle avoidance and path-planning.

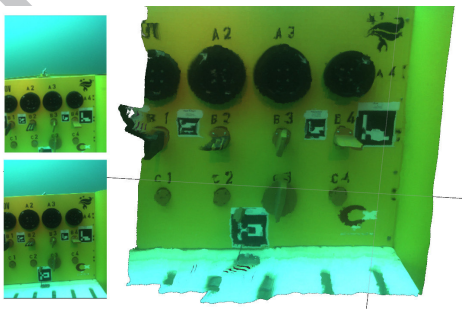
Furthermore, we provide here a quantitative analysis of the stereo system. The perception error depends of course on multiple factors. First of all, there is the accuracy of



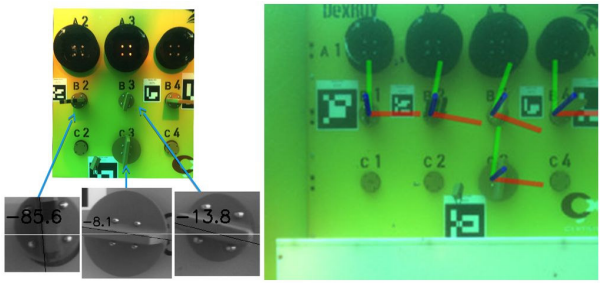
(a) Image enhancement to compensate turbidity



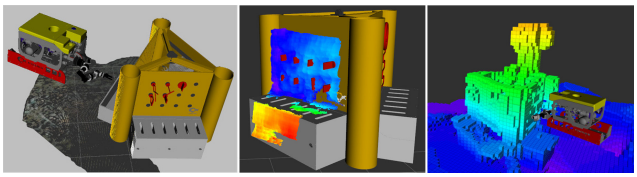
(b) Disparity maps for sparse (center) and dense stereo (right) of the DexROV panel (left)



(c) RGBD point cloud (right) from dense stereo



(d) Localization and state detection of valve handles



(e) Mixed reality display of different forms of maps: RBD point cloud (left), dense disparity (center), octree (right)

Figure 17: Qualitative examples of underwater vision results using the presented stereo system.

Figure 18: The Jacobs stereo system was among others tested in two extensive field trials off the shore of Marseille. The camera system was integrated on a commercial Apache ROV (left), which is deployed from the COMEX Janus II vessel (center). A mockup panel structure is used to test different application scenarios like operations in oil- and gas-production (right).

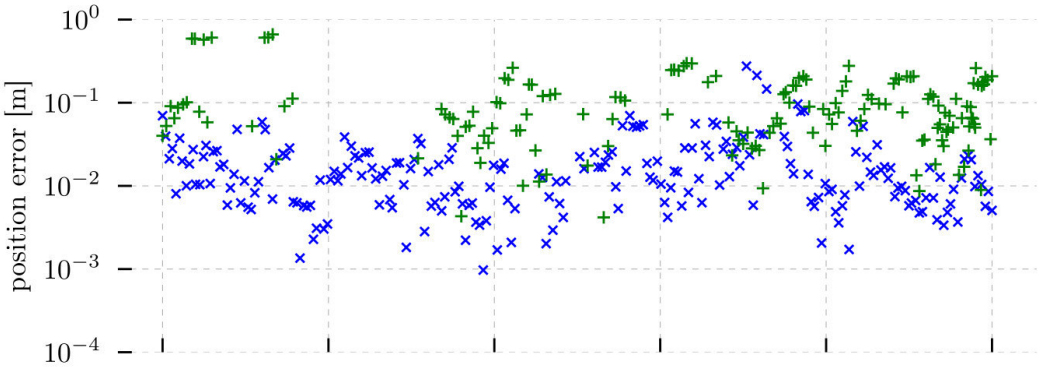


Figure 19: Residual errors and of panel marker detection during a mission for simulated noise-free (blue 'x') and field-trial underwater (green '+') conditions.

the stereo system itself, which also includes the camera calibration and the image rectification aspects. Also, there are the influences of the underwater visibility conditions, which lead to noise and image degradation. Last but not least, there is the influence of the accuracy of the vision methods used for the computation of metric information from the stereo data.

The DexROV panel is augmented with ArUco Augmented Reality (AR) markers, which can be reliably detected by computer vision (Fig.20). The AR-markers are placed at known locations on the panel. This can be used to estimate the pose of the panel relative to the stereo system, respectively to the ROV. Especially when two or more AR-markers are perceived, a residual error can be determined. Fig.19 shows the residual errors in marker detection during a real-world mission (blue). The average error of the position estimation of the panel in the real-world experiments is 11.8cm (±12.6 cm) while the vehicle is operating in a distance range of 2.5m to 6m to the panel where multiple markers can be detected.

Furthermore, the errors during play-back of the mission in a high-fidelity simulation under completely noise-free conditions are included for comparison (green). I.e., the simulated vehicle follows the same trajectory as the real one and the vision system is fed with images from a 3D simulation of the environment. The effects of underwater image formation, e.g., haze and noise, can be incorporated in the simulation, but they are not used here. The noise-free, simulated perception hence provides a kind of best case baseline. It can be observed that the real-world performance (blue) is not too far off this theoretical best case (green).

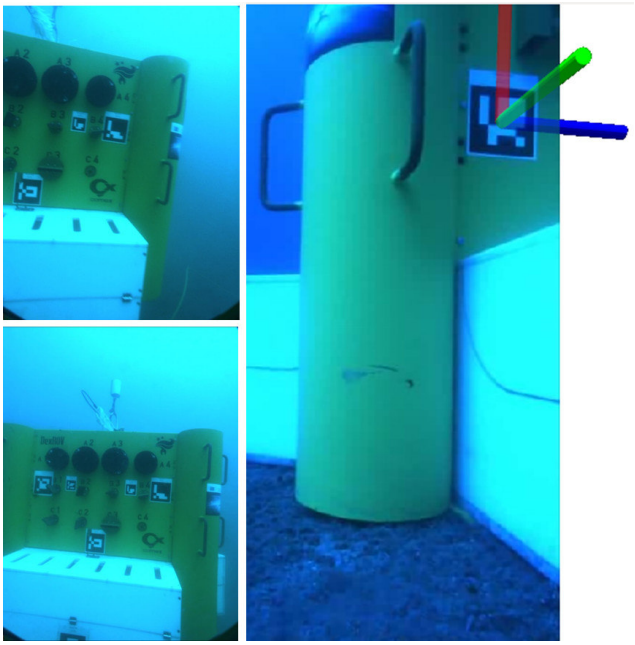


Figure 20: The DexROV mock-up panel is equipped with multiple AR-markers at known locations, i.e., with ground truth relative distances between each other (left). The markers can be detected and localized using computer vision (right).

6. Conclusions

In this article, an intelligent stereo vision system for deep-sea operations was presented, which was developed with model based design. A formalized approach to the component selection for the stereo set-up was introduced, which is of general interest for designing stereo systems. Furthermore, the concrete aspects of deep sea operations were addressed, especially the design and optimization of the camera housings by a Finite Element Method (FEM). Last but not least, results from experiments were presented including a validation of the stereo performance in air, robustness tests of the bottles in pressure tanks, and field trials of the complete system off the shore of Marseille on a commercial Remotely Operated Vehicle (ROV).

Acknowledgements

The research leading to the presented results was supported in part by the European Community's Horizon2020 program under grant agreement n. 635491 "Dexterous ROV: effective dexterous ROV operations in presence of communication latencies (DexROV)".

[1] G. N. Desouza and A. C. Kak, "Vision for mobile robot navigation: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 2, pp. 237–267, Feb 2002.

[2] P. Cieslak, P. Ridao, and M. Giergiel, "Autonomous underwater panel operation by girona500 uvms: A practical approach to autonomous underwater manipulation," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, May 2015, pp. 529–536.

[3] D. Lodi Rizzini, F. Kallasi, J. Aleotti, F. Oleari, and S. Caselli, "Integration of a stereo vision system into an autonomous underwater vehicle for pipe manipulation tasks," *Computers and Electrical Engineering*, vol. 58, pp. 560–571, 2017.

[4] Z. Zhang, H. Rebecq, C. Forster, and D. Scaramuzza, "Benefit of large field-of-view cameras for visual odometry," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, May 2016, pp. 801–808.

[5] J. Chen, S. Khatibi, and W. Kulesza, "Planning of a multi stereo visual sensor system - depth accuracy and variable baseline approach," in *2007 3DTV Conference*, May 2007, pp. 1–4.

[6] D. C. Herath, K. R. S. Kodagoda, and G. Dissanayake, "Modeling errors in small baseline stereo for slam," in *2006 9th International Conference on Control, Automation, Robotics and Vision*, Dec 2006, pp. 1–6.

[7] D. Pojar, P. Jeong, and S. Nedevschi, "Robust visual odometry using stereo reconstruction error model," in *Intelligent Computer Communication and Processing (ICCP), 2012 IEEE International Conference on*, Aug 2012, pp. 149–154.

[8] G. Dubbelman and F. C. A. Groen, "Bias reduction for stereo based motion estimation with applications to large scale visual odometry," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, June 2009, pp. 2222–2229.

[9] L. Matthies and S. Shafer, "Error modeling in stereo navigation," *IEEE Journal on Robotics and Automation*, vol. 3, no. 3, pp. 239–248, June 1987.

[10] J. J. Rodriguez and J. K. Aggarwal, "Stochastic analysis of stereo quantization error," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 5, pp. 467–470, May 1990.

[11] T. Zhang and T. Boult, "Realistic stereo error models and finite optimal stereo baselines," in *Applications of Computer Vision (WACV), 2011 IEEE Workshop on*, Jan 2011, pp. 426–433.

[12] N. V. Srinivasulu, S. Khan, and S. Jaikrishna, "Design and analysis of submarine radome," in *Proceedings of the International Conference on Research and Innovations in Mechanical Engineering*, 01 2014, pp. 11–26.

[13] T.-H. Joung, I. Nho, and P.-M-Lee, "A study on the structural design and analysis of unmanned underwater vehicle," *International Society of Offshore and Polar Engineers*, no. 15, June 2005.

[14] A. Yousefpour, "Design, analysis, manufacture, and test of composite pressure vessels and finite element analysis of metallic frame for deep ocean underwater vehicle applications," 12 2017.

[15] K. K. Smitha and C. G. Nandakumar, "Finite element analysis of autonomous underwater glider," *IJAERD*, vol. 3, no. 5, pp. 1–6, December 2013.

[16] K. Muljowidodo, S. A. Nugroho, N. Prayogo, B. Gunadharma, and A. Budiyono, "Design and analysis of modular composites drybox hull of hybrid autonomous underwater vehicle," *Indian Journal of Geo-Marine Sciences*, vol. 41, no. 6, pp. 563–568, December 2012.

[17] A. Birk, T. Doernbach, C. Mueller, T. Luczynski, A. G. Chavez, D. Köhntopp, A. Kupcsik, S. Calinon, A. Tanwani, G. Antonelli, P. d. Lillo, E. Simetti, G. Casalino, G. Indiveri, L. Ostuni, A. Turetta, A. Caffaz, P. Weiss, T. Gobert, B. Chemisky, J. Gancet, T. Siedel, S. Govindaraj, X. Martinez, and P. Letier, "Dexterous underwater manipulation from distant onshore locations," *IEEE Robotics and Automation Magazine (RAM)*, vol. 25, no. 4, 2018.

[18] T. Treibitz, Y. Y. Schechner, and H. Singh, "Flat refractive geometry," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008, pp. 1–8.

[19] A. Sedlazeck and R. Koch, "Calibration of Housing Parameters for Underwater Stereo-Camera Rigs," in *Proceedings of the British Machine Vision Conference*, J. Hoey, Ed. BMVA Press, 2011.

[20] T. Treibitz, Y. Schechner, C. Kunz, and H. Singh, "Flat refractive geometry," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 1, pp. 51–65, 2012.

[21] T. Dolereit, U. F. v. Lukas, and A. Kuijper, "Underwater stereo

calibration utilizing virtual object points," in *OCEANS*. IEEE, 2015, pp. 1–7.

[22] T. Luczynski, M. Pfingsthorn, and A. Birk, "The pinax-model for accurate and efficient refraction correction of underwater cameras in flat-pane housings," *Ocean Engineering*, vol. 133, pp. 9–22, 2017.

[23] M. Welsch, "Bewertung von spannungsspitzen und singularitäten in fem-rechnungen," in *Norddeutsches Simulationsforum, Hamburg*, vol. 11, 2015.

[24] T. Doernbach, A. G. Chavez, C. A. Mueller, and A. Birk, "High-fidelity deep-sea perception using simulation in the loop," in *IFAC Conference on Control Applications in Marine Systems (CAMS)*, 2018.

[25] C. A. Mueller, T. Doernbach, A. G. Chavez, D. Köhntopp, and A. Birk, "Robust continuous system integration for critical deep-sea robot operations using knowledge-enabled simulation in the loop," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018.

[26] T. Luczynski, T. Fromm, S. Govindaraj, C. A. Mueller, and A. Birk, "3d grid map transmission for underwater mapping and visualization under bandwidth constraints," in *IEEE Oceans*. IEEE press, 2017.

[27] T. Luczynski and A. Birk, "Underwater image haze removal with an underwater-ready dark channel prior," in *IEEE Oceans*. IEEE press, 2017.

Highlights

- An algorithm for stereo vision hardware and setup parameters selection is proposed
- Model based design of the deep-sea housings is presented
- Proposed algorithm is evaluated in the experiment
- Complete system is validated in the deep-sea operation